BOUNDED LENGTH INTERVALS CONTAINING TWO PRIMES AND AN ALMOST-PRIME

JAMES MAYNARD

Abstract. Goldston, Pintz and Yıldırım have shown that if the primes have 'level of distribution' θ for some $\theta > 1/2$ then there exists a constant $C(\theta)$, such that there are infinitely many integers n for which the interval $[n, n + C(\theta)]$ contains two primes. We show under the same assumption that for any integer $k \geq 1$ there exists constants $D(\theta, k)$ and $r(\theta, k)$, such that there are infinitely many integers n for which the interval $[n, n+D(\theta, k)]$ contains two primes and k almost-primes, with all of the almost-primes having at most $r(\theta, k)$ prime factors. If θ can be taken as large as 0.99, and provided that numbers with 2, 3, or 4 prime factors also have level of distribution 0.99, we show that there are infinitely many integers n such that the interval $[n, n + 90]$ contains 2 primes and an almost-prime with at most 4 prime factors.

1. Introduction

We are interested in trying to understand how small gaps between primes can be. If we let p_n denote the n^{th} prime, it is conjectured that

$$\liminf_n p_{n+1} - p_n = 2. \tag{1.1}$$

This is the famous twin prime conjecture. Unfortunately we appear unable to prove any results of this strength. The best unconditional result is due to Goldston, Pintz and Yıldırım [5] which states that

$$\liminf_n \frac{p_{n+1} - p_n}{\sqrt{\log p_n}(\log\log p_n)^2} < \infty. \tag{1.2}$$

Therefore we do not know that $\liminf p_{n+1} - p_n$ is finite.

The method of [5] relies heavily on results about primes in arithmetic progressions. We say that the primes have 'level of distribution' θ if for any constant A there is a constant $C = C(A)$ such that

$$\sum_{q \leq x^\theta (\log x)^{-C}} \max_{\substack{a \\ (a,q)=1}} \left| \sum_{\substack{p \equiv a \pmod q \\ p \leq x}} 1 - \frac{\mathrm{Li}(x)}{\phi(q)} \right| \ll_A \frac{x}{(\log x)^A}. \tag{1.3}$$

The Bombieri-Vinogradov theorem states that the primes have level of distribution $1/2$, and this is a major ingredient in the proof of the Goldston-Pintz-Yıldırım result.

2010 *Mathematics Subject Classification.* 11N05, 11N35, 11N36.

Supported by EPSRC Doctoral Training Grant EP/P505216/1 .

If we could improve the Bombieri-Vinogradov theorem to show that the primes have level of distribution θ for some constant $\theta > 1/2$, then it would follow from [4][Theorem 1] that there is a constant $D = D(\theta)$ such that

$$\liminf_n p_{n+1} - p_n < D, \tag{1.4}$$

and so there would be infinitely many bounded gaps between primes. It is believed that such improvements to the Bombieri-Vinogradov theorem are true, and Elliott and Halberstam [1] conjectured the following much stronger result.

Conjecture (Elliott-Halberstam Conjecture)**.** *For any fixed $\epsilon > 0$, the primes have level of distribution $1 - \epsilon$.*

Friedlander and Granville [2] have shown that the primes do not have level of distribution 1, and so the Elliott-Halberstam conjecture represents the strongest possible result of this type.

Under the Elliott-Halberstam conjecture the Goldston-Pintz-Yıldırım method gives [4] that

$$\liminf_n p_{n+1} - p_n \leq 16. \tag{1.5}$$

If we consider the length of 3 or more consecutive primes, however, we are unable to prove as strong results, even under the full strength of the Elliott-Halberstam conjecture. In particular we are unable to prove that there are infinitely many intervals of bounded length that contain at least 3 primes. The Goldston-Pintz-Yıldırım methods can still be used, but even with the Elliot-Halberstam conjecture we are only able to prove that

$$\liminf_n \frac{p_{n+2} - p_n}{\log p_n} = 0. \tag{1.6}$$

This should be contrasted with the following conjecture.

Conjecture (Prime k-tuples conjecture)**.** *Let $\mathcal{L} = \{L_1, \dots, L_k\}$ be a set of integer linear functions whose product has no fixed prime divisor. Then there are infinitely many n for which all of $L_1(n), L_2(n), \dots, L_k(n)$ are simultaneously prime.*

By 'no fixed prime divisor' above we mean that for every prime p there is an integer n_p such that $L_i(n_p)$ is coprime to p for all $1 \leq i \leq k$. We call such a set of linear functions *admissible*.

We note that $\{n, n+2, n+6\}$ is an admissible set of linear functions, and so the prime k-tuples conjecture predicts that $\liminf_n p_{n+2} - p_n \leq 6$ (it is easy to verify that one cannot have $p_{n+2} - p_n < 6$ for $n > 2$). More generally, for any constant $k > 0$ the conjecture predicts that $\liminf p_{n+k} - p_n < \infty$, and so there are infinitely many intervals of bounded size containing at least k primes.

At the moment the prime k-tuples conjecture appears beyond the techniques currently available to us. As an approximation to the conjecture, it is common to look for *almost-prime* numbers instead of primes, where almost-prime indicates that the number has only a 'few' prime factors.

Graham, Goldston, Pintz and Yıldırım [3] have shown that given an integer k, there are infinitely many intervals of bounded length (depending on k) containing at least k integers each with exactly two prime factors. It is a classical result of Halberstam and Richert [6] that there are infinitely many intervals of bounded length (depending on k) which contain

a prime and at least k numbers each with at most r prime factors for r sufficiently large (depending on k).

We investigate, under the assumption that the primes have level of distribution $\theta > 1/2$, whether there are infinitely many intervals of bounded length (depending on k) containing 2 primes and k numbers each with at most r prime factors.

2. Initial Hypotheses

We will work with an assumption either on the distribution of primes in arithmetic progressions of level θ, or a stronger assumption on numbers with exactly r prime factors each of which is of a given size.

Given constants $0 \le \eta_i \le \delta_i \le 1$ for $1 \le i \le r$ we define

$$\beta_{r,\eta,\delta}(n) = \begin{cases} 1, & n = p_1 p_2 \dots p_r \text{ with } n^{\eta_i} \le p_i \le n^{\delta_i} \text{ for } 1 \le i \le r, \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

We put

$$\Delta(x; q, a) = \sum_{\substack{x < p \le 2x \\ p \equiv a \pmod q}} 1 - \frac{1}{\phi(q)} \sum_{x < p \le 2x} 1, \tag{2.2}$$

$$\Delta_{r,\eta,\delta}(x; q, a) = \sum_{\substack{x < p \le 2x \\ p \equiv a \pmod q}} \beta_{r,\eta,\delta}(n) - \frac{1}{\phi(q)} \sum_{x < p \le 2x} \beta_{r,\eta,\delta}(n), \tag{2.3}$$

$$\Delta^*(x; q) = \max_{y \le x} \max_{\substack{a \\ (a,q)=1}} |\Delta(y; q, a)|, \tag{2.4}$$

$$\Delta^*_{r,\eta,\delta}(x; q) = \max_{y \le x} \max_{\substack{a \\ (a,q)=1}} \left|\Delta_{r,\eta,\delta}(y; q, a)\right|. \tag{2.5}$$

We can now state the two hypotheses that we will consider, the Bombieri-Vinogradov hypothesis of level θ, BV(θ), and the generalised Bombieri-Vinogradov hypothesis of level θ for E_r numbers, GBV(θ,r).

Hypothesis BV(θ). *For every constant $A > 0$ and integer $h > 0$ there is a constant $C = C(A, h)$ such that if $Q \le x^\theta (\log x)^{-C}$ then we have*

$$\sum_{q \le Q} \mu^2(q) h^{\omega(q)} \Delta^*(x; q) \ll_A x(\log x)^{-A}.$$

Hypothesis GBV(θ,r). *For every constant $A > 0$ and integer $h > 0$ there is a constant $C = C(A, h)$ such that if $Q \le x^\theta (\log x)^{-C}$ then uniformly for $0 \le \eta_i \le \delta_i \le 1$ $(1 \le i \le r)$ we have*

$$\sum_{q \le Q} \mu^2(q) h^{\omega(q)} \Delta^*_{r,\eta,\delta}(x; q) \ll_A x(\log x)^{-A}.$$

We note that by standard arguments in sieve methods (see, for example, [6][Lemma 3.5]) Hypothesis BV(θ) follows from the primes having level of distribution θ.

3. Statement of Results

Theorem 3.1. *Let $k \geq 1$ be an integer. Let $1/2 < \theta < 0.99$. Assume Hypothesis BV(θ) holds. Let*

$$r = \frac{240k^2}{(2\theta - 1)^3}.$$

Then there are infinitely many integers n such that the interval $[n, n + C(k, \theta)]$ contains two primes and k integers, each with at most r prime factors.

Theorem 3.2. *Let $\theta \geq 0.99$, and assume Hypothesis GBV(θ,r) holds for $1 \leq r \leq 4$. Then there exist infinitely many integers n such that the interval $[n, n + 90]$ contains two primes and one other integer with at most 4 prime factors.*

4. Proof of Theorem 3.1

We consider two finite disjoint sets of integer linear functions $\mathcal{L}_1^{(1)} = \{L_1^{(1)}, \dots, L_{k_1}^{(1)}\}$ and $\mathcal{L}_2^{(1)} = \{L_{k_1+1}^{(1)}, \dots, L_{k_1+k_2}^{(1)}\}$, whose union $\mathcal{L}^{(1)} = \mathcal{L}_1^{(1)} \cup \mathcal{L}_2^{(1)}$ is admissible. (We recall that a such set is admissible if for every prime p there is an integer n_p such that every function evaluated at n_p is coprime to p).

We wish to show that there are infinitely many n for which two of the functions from $\mathcal{L}_1^{(1)}$ take prime values at n, and at all of the functions from $\mathcal{L}^{(2)}$ take almost-prime values at n.

Since we are only interested in showing there are infinitely many such n, we adopt a normalisation of our linear functions, as done originally by Heath-Brown [7] which simplifies our argument. By considering $L_i(n) = L_i^{(1)}(An + B)$ for suitable constants A and B we may assume that the functions L_i satisfy the following conditions.

(1) The functions $L_i(n) = a_i n + b_i$ $(1 \leq i \leq k_1 + k_2)$ are distinct with $a_i > 0$.
(2) Each of the coefficients a_i is composed of the same primes none of which divides the b_j.
(3) If $i \neq j$, then any prime factor of $a_i b_j - a_j b_i$ divides each of the a_l.

We let $\mathcal{L}_1 = \{L_1, \dots, L_{k_1}\}$ and $\mathcal{L}_2 = \{L_{k_1+1}, \dots, L_{k_1+k_2}\}$.

We now consider the sum

$$S(N; \mathcal{L}_1, \mathcal{L}_2) = \sum_{N \leq n \leq 2N} \left(\sum_{L \in \mathcal{L}_1} \chi_1(L(n)) + \sum_{L \in \mathcal{L}_2} \chi_r(L(n)) - k_2 - 1 \right) \left(\sum_{\substack{d | \Pi(n) \\ d \leq R}} \lambda_d \right)^2, \tag{4.1}$$

where

$$\chi_r(n) = \begin{cases} 1, & n \text{ has at most } r \text{ prime factors} \\ 0, & \text{otherwise,} \end{cases} \tag{4.2}$$

$$R = N^{\theta/2} (\log N)^{-C}, \tag{4.3}$$

$$\Pi(n) = \prod_{L \in \mathcal{L}_1 \cup \mathcal{L}_2} L(n), \tag{4.4}$$

and the λ_d are real numbers which we declare later. $C > 0$ is a constant chosen sufficiently large so we can use the estimates of hypotheses $BV(\theta)$ or $GBV(\theta, r)$.

If we can show that $S > 0$ then we know there must be at least one $n \in [N, 2N]$ for which the terms in parentheses give a positive contribution to S. The second term in our expression for S is a square, and so is always non-negative. We see that the first term in parentheses is positive only when there are at least two primes and k_2 numbers each with at most r prime factors amongst the $L_i(n)$ $(1 \leq i \leq k_1 + k_2)$. If we choose all our original functions to be of the form $L_i^{(1)}(n) = n + h_i$ (with $h_i \geq 0$) then all these integers then lie in an interval $[m, m + H]$, where $H = \max_i h_i$.

Thus it is sufficient to show that $S > 0$ for any large N to prove Theorem 3.1. We can get such a bound by following a method similar to Goldston, Pintz and Yıldırım [5], which we refer to as the GPY method.

To simplify notation we put

$$\Lambda^2(n) = \left(\sum_{\substack{d \mid \Pi(n) \\ d \leq R}} \lambda_d \right)^2. \tag{4.5}$$

To avoid confusion we mention that $\Lambda^2(n)$ is unrelated to the Von-Mangold function.

We expect to be able to show that $S > 0$ for suitably large k_1 and r (depending on k_2) when the primes have level of distribution $\theta > 1/2$. This is because the original GPY method shows that for sufficiently large size of k_1 (depending on k_2 and ϵ) we can choose the λ_d to give

$$\sum_{N \leq n \leq 2N} \sum_{L \in \mathcal{L}_1} \chi_1(L(n)) \Lambda^2(n) \geq (2\theta - \epsilon) \sum_{N \leq n \leq 2N} \Lambda^2(n). \tag{4.6}$$

Moreover, since $\Lambda^2(n)$ is small when $\Pi(n)$ has many prime factors, we expect for sufficiently large r (depending on k_2 and ϵ) that

$$\sum_{N \leq n \leq 2N} \sum_{L \in \mathcal{L}_2} (1 - \chi_r(L(n))) \Lambda^2(n) \leq \epsilon \sum_{N \leq n \leq 2N} \Lambda^2(n). \tag{4.7}$$

And so provided that $\theta > 1/2 + \epsilon$ we expect that

$$S \gg_\epsilon \sum_{N \leq n \leq 2N} \Lambda^2(n) > 0. \tag{4.8}$$

Although the method of Graham, Goldston, Pintz and Yıldırım allows one to estimate similar sums involving numbers with a fixed number of prime factors, these results rely on level-of-distribution results for such numbers, which we are not assuming in Theorem 3.1. Instead we proceed by noting that any integer which is at most $2N$ and has more than r prime factors must have a prime factor of size at most $(2N)^{1/(r+1)}$. Thus for $n \leq 2N$

$$\chi_r(n) \geq 1 - \sum_{\substack{p \mid n \\ p \leq (2N)^{1/(r+1)}}} 1. \tag{4.9}$$

Substituting this into our expression for S we have

$$S \geq \sum_{N\leq n\leq 2N}\left(\sum_{L\in\mathcal{L}_1}\chi_1(L(n)) - 1 - \sum_{L\in\mathcal{L}_2}\sum_{\substack{p|L(n)\\ p\leq(2N)^{1/r+1}}} 1\right)\Lambda^2(n)$$

$$= \sum_{L\in\mathcal{L}_1} Q_1(L) - Q_2 - \sum_{L\in\mathcal{L}_2} Q_3(L), \tag{4.10}$$

where

$$Q_1(L) = \sum_{N\leq n\leq 2N}\chi_1(L(n))\Lambda^2(n), \tag{4.11}$$

$$Q_2 = \sum_{N\leq n\leq 2N}\Lambda^2(n), \tag{4.12}$$

$$Q_3(L) = \sum_{N\leq n\leq 2N}\sum_{\substack{p|L(n)\\ p\leq(2N)^{1/(r+1)}}}\Lambda^2(n). \tag{4.13}$$

The choice of good values for λ_d and the corresponding evaluation of Q_1, Q_2, Q_3 already exists in the literature. We quote from [8][Proposition 4.1] taking $W_0(t) = 1$ and[3][Theorem 7 and Theorem 9]. We note that [3][Theorem 9] does not require that E_2-numbers have level of distribution θ, so Hypothesis $BV(\theta)$ is sufficient for the statement to hold. These results give for a fixed polynomial P, for $k = k_1 + k_2 = \#(\mathcal{L}_1 \cup \mathcal{L}_2)$, for $L \in \mathcal{L}$ and for sufficiently large C that we can choose the λ_d such that

$$Q_1(L) \sim \frac{\mathfrak{S}(\mathcal{L})N(\log R)^{k+1}}{(k-2)!\log N}\int_0^1 \tilde{P}(1-t)^2 t^{k-2}dt, \tag{4.14}$$

$$Q_2 \sim \frac{\mathfrak{S}(\mathcal{L})N(\log R)^{k}}{(k-1)!}\int_0^1 P(1-t)^2 t^{k-1}dt, \tag{4.15}$$

$$Q_3(L) \sim \frac{\mathfrak{S}(\mathcal{L})N(\log R)^{k}}{(k-1)!}I, \tag{4.16}$$

where

$$\mathfrak{S}(\mathcal{L}) \text{ is a positive constant depending only on } \mathcal{L}, \tag{4.17}$$

$$\tilde{P}(x) = \int_0^x P(t)dt, \tag{4.18}$$

$$I = \int_0^\delta \frac{1}{y}\int_{1-y}^1 P(1-t)^2 t^{k-1}dtdy$$

$$+ \int_0^\delta \frac{1}{y}\int_0^{1-y} \left(P(1-t) - P(1-t-y)\right)^2 t^{k-1}dtdy, \tag{4.19}$$

$$\delta = \frac{2}{\theta(r+1)}. \tag{4.20}$$

Here the asymptotic for Q_3 is valid only for $R^2(2N)^{1/(r+1)} \leq N(\log N)^{-C}$, and so we introduce the condition

$$r+1 > \frac{1}{1-\theta} \tag{4.21}$$

to ensure that this is satisfied for N sufficiently large. All the other asymptotics are valid without further conditions.

We choose the polynomial

$$P(x) = x^l, \tag{4.22}$$

where $l \geq 0$ is an integer to be declared later. To ease notation we let

$$C(\mathcal{L}) = \frac{\mathfrak{S}(\mathcal{L})N(\log R)^k(2l)!}{(k+2l)!}. \tag{4.23}$$

We note that for $a, b \in \mathbb{N}$

$$\int_0^1 x^a(1-x)^b dx = \frac{a!b!}{(a+b+1)!}. \tag{4.24}$$

Thus we see that

$$\sum_{L\in\mathcal{L}_1} Q_1(L) \sim \theta\left(\frac{2l+1}{l+1}\right)\left(\frac{k_1}{k+2l+1}\right)C(\mathcal{L}), \tag{4.25}$$

$$Q_2 \sim C(\mathcal{L}). \tag{4.26}$$

We follow a similar approach to Graham, Goldston, Pintz and Yıldırım [3] to estimate Q_3. We let

$$I = \int_0^\delta \frac{F(y)}{y}dy, \tag{4.27}$$

where

$$\begin{aligned} F(y) &= \int_{1-y}^1 P(1-t)^2 t^{k-1}dt + \int_0^{1-y} (P(1-t) - P(1-t-y))^2\, t^{k-1}dt \\ &= \int_0^1 P(1-t)^2 t^{k-1}dt + \int_0^{1-y} P(1-t-y)^2 t^{k-1}dt \\ &\quad - 2\int_0^{1-y} P(1-t)P(1-t-y)t^{k-1}dt. \end{aligned} \tag{4.28}$$

We recall that $P(x) = x^l$, and note that

$$P(1-t) = (1-t)^l = (1-t-y)^l + \sum_{j=1}^{l}\binom{l}{j}y^j(1-t-y)^{l-j}. \tag{4.29}$$

Thus

$$\begin{aligned} F(y) &= \int_0^1 (1-t)^{2l}t^{k-1}dt + \int_0^{1-y}(1-t-y)^{2l}t^{k-1}dt \\ &\quad - 2\int_0^{1-y}(1-t-y)^{2l}t^{k-1}dt - 2\sum_{j=1}^{l}\binom{l}{j}y^j\int_0^{1-y}(1-t-y)^{2l-j}t^{k-1}dt \\ &\leq \int_0^1 (1-t)^{2l}t^{k-1}dt - \int_0^{1-y}(1-t-y)^{2l}t^{k-1}dt \\ &= \frac{(k-1)!(2l)!}{(k+2l)!}\left(1-(1-y)^{k+2l}\right). \end{aligned} \tag{4.30}$$

Substituting this into (4.27) gives

$$I = \int_0^\delta \frac{F(y)}{y}dy \leq \frac{(k-1)!(2l)!}{(k+2l)!}\int_0^\delta \frac{1-(1-y)^{k+2l}}{y}dy \tag{4.31}$$

$$= \frac{(k-1)!(2l)!}{(k+2l)!}\sum_{j=0}^{k+2l-1}\int_0^\delta (1-y)^j dy \tag{4.32}$$

$$\leq \frac{(k-1)!(2l)!}{(k+2l)!}\delta(k+2l). \tag{4.33}$$

Therefore from (4.16), (4.23) and (4.33), for any $L \in \mathcal{L}_2$ we have that

$$Q_3(L) \leq C(\mathcal{L})\left(\delta(k+2l)+o(1)\right). \tag{4.34}$$

Substituting (4.25), (4.26) and (4.34) into (4.10) we see

$$S \geq C(\mathcal{L})\left(\theta\left(\frac{2l+1}{l+1}\right)\left(\frac{k_1}{k+2l+1}\right)-1-k_2\delta(k+2l)+o(1)\right). \tag{4.35}$$

We let

$$k+2l+1 = \lceil C_1(2\theta-1)^{-2}\rceil, \qquad l+1 = \lceil C_2(2\theta-1)^{-1}\rceil \tag{4.36}$$

for some C_1, C_2. This gives

$$\begin{aligned}\frac{S}{C(\mathcal{L})} &\geq \theta\left(2-\frac{2\theta-1}{C_2}\right)\left(1-\frac{2C_2(2\theta-1)}{C_1}-\frac{(k_2+1)(2\theta-1)^2}{C_1}\right)-1\\ &\qquad -k_2\delta C_1(2\theta-1)^{-2}+o(1)\\ &=(2\theta-1)\left(1-\frac{\theta}{C_2}-\frac{4\theta C_2}{C_1}-\frac{2k_2(2\theta-1)\theta}{C_1}+\frac{(1+k_2)\theta(2\theta-1)^2}{C_1C_2}\right)\\ &\qquad -k_2\delta C_1(2\theta-1)^{-2}+o(1).\end{aligned} \tag{4.37}$$

We let

$$C_1 = 40k_2, \qquad C_2 = 3. \tag{4.38}$$

We see from (4.36) that this choice of C_1 and C_2 corresponds to positive integer values for k_1 and l for any choice of $0.5 < \theta \leq 0.99$ or k_2, and so is a valid choice.

Since k_2 is a positive integer and $1/2 < \theta \leq 1$, this gives

$$\begin{aligned}\frac{S}{C(\mathcal{L})} &\geq (2\theta-1)\left(\frac{30-17\theta-5\theta^2+2\theta^3}{30}\right)-40k_2^2(2\theta-1)^{-2}\delta+o(1)\\ &\geq (2\theta-1)\left(\frac{31-21\theta}{30}\right)-40k_2^2(2\theta-1)^{-2}\delta+o(1).\end{aligned} \tag{4.39}$$

Thus $S > 0$ for large N if δ is chosen such that

$$\delta < \frac{(2\theta-1)^3(31-21\theta)}{1200k_2^2}. \tag{4.40}$$

We recall $\delta = 2/\theta(r+1)$, so S is positive provided r is chosen larger than

$$\frac{2400k_2^2}{(2\theta-1)^3\theta(31-21\theta)}-1 < \frac{240k_2^2}{(2\theta-1)^3}. \tag{4.41}$$

We note that if $r = 240k_2^2/(2\theta-1)^3$ then for $\theta < 0.99$ the condition (4.21) is satisfied. This completes the proof of Theorem 3.1.

We remark that by choosing $L(n) = n + h$ with $h < H$ for $L \in \mathcal{L}_2$ and $L(n) = n + h$ with $h > H$ for $L \in \mathcal{L}_1$ we can ensure that of the $k_2 + 2$ almost-primes we find, the largest two are primes.

5. Proof of Theorem 3.2

We can get better quantitative results for the number of prime factors involved in our almost-prime if we assume a fixed level of distribution result for almost-primes and for primes, and then follow the work of [3].

We consider the same sum S, but now we assume that $\mathcal{L}_2 = \{L_0\}$ and we take $r = 4$. Thus $k_2 = 1$ and $k = k_1 + 1$.

$$S = S(N; \mathcal{L}_1, \{h_0\}) = \sum_{N \le n \le 2N} \left(\sum_{L \in \mathcal{L}_1} \chi_1(L(n)) + \chi_4(L_0(n)) - 2 \right) \Bigg(\sum_{\substack{d | \Pi(n) \\ d \le R}} \lambda_d \Bigg)^2$$
$$= \sum_{L \in \mathcal{L}_1} Q_1(L) + Q_1'(L_0) - Q_2, \tag{5.1}$$

where $Q_1(L)$, Q_2 are as before and

$$Q_1' = \sum_{N \le n \le 2N} \chi_4(L_0(n)) \Bigg(\sum_{\substack{d | \Pi(n) \\ d \le R}} \lambda_d \Bigg)^2 \tag{5.2}$$

$$R = N^{0.99/2} (\log N)^C. \tag{5.3}$$

As before, C is a suitably large positive constant.

We split the contribution to Q_1' depending on whether $L_0(n)$ has exactly 1, 2, 3 or 4 prime factors. Thus

$$Q_1' = Q_1(L_0) + Q_{12}' + Q_{13}' + Q_{14}', \tag{5.4}$$

where

$$Q_{1j}' = \sum_{N \le n \le 2N} \beta_j(L_0(n)) \Bigg(\sum_{\substack{d | \Pi(n) \\ d \le R}} \lambda_d \Bigg)^2, \tag{5.5}$$

and

$$\beta_j(n) = \begin{cases} 1, & n \text{ has exactly } j \text{ prime factors} \\ 0, & \text{otherwise.} \end{cases} \tag{5.6}$$

For technical reasons we find it harder to deal with terms arising when $L(n)$ has a prime factor less than N^ϵ or no prime factor greater than $N^{1/2}$. Thus we obtain a lower bound for Q_{1j}' by replacing $\beta_j(L_0(n))$ with $\beta_j'(L_0(n))$, where

$$\beta_j'(n) = \begin{cases} 1, & n = p_1 p_2 \dots p_j \text{ with } n^\epsilon < p_1 < \cdots < p_j \text{ and } n^{0.505} < p_j \\ 0, & \text{otherwise.} \end{cases} \tag{5.7}$$

We can then obtain these asymptotic lower bounds. By following an equivalent argument to [9] and [8][Proposition 4.2] but using Hypothesis GBV(0.99, j) to bound the error terms we have

$$Q'_{1j} \geq (1+o(1))\frac{\mathfrak{S}(\mathcal{L})(\log R)^{k+1}}{(k-2)!\log N}J_j, \tag{5.8}$$

where

$$J_r = \int_{(x_1,\dots,x_{r-1})\in\mathcal{A}_r}\frac{I_1(Bx_1,\dots,Bx_{r-1})}{\left(\prod_{i=1}^{r-1}x_i\right)\left(1-\sum_{i=1}^{r-1}x_i\right)}dx_1\dots dx_{r-1}, \tag{5.9}$$

$$I_1 = \int_0^1\left(\sum_{J\subset\{1,\dots,r-1\}}(-1)^{|J|}\tilde{P}^+(1-t-\sum_{i\in J}x_i)\right)^2 t^{k-2}dt, \tag{5.10}$$

$$\tilde{P}^+(x) = \begin{cases}\int_0^x P(t)dt, & x\geq 0\\ 0, & \text{otherwise,}\end{cases} \tag{5.11}$$

$$B = \frac{2}{0.99}, \tag{5.12}$$

$$\mathcal{A}_r = \left\{x\in[0,1]^{r-1} : \epsilon < x_1 < \dots < x_{r-1}, \sum_{i=1}^{r-1}x_i < B^{-1}\right\}. \tag{5.13}$$

As before, by Hypothesis BV(0.99) we also have for any $L\in\mathcal{L}$

$$Q_1(L) \sim \frac{\mathfrak{S}(\mathcal{L})(\log R)^{k+1}}{(k-2)!\log N}\int_0^1\tilde{P}(1-t)^2t^{k-2}dt, \tag{5.14}$$

$$Q_2 \sim \frac{\mathfrak{S}(\mathcal{L})(\log R)^{k}}{(k-1)!}\int_0^1 P(1-t)^2t^{k-1}dt. \tag{5.15}$$

Thus we have that

$$S \geq \frac{\mathfrak{S}(\mathcal{L})(\log R)^{k}}{(k-1)!}\left(\frac{0.99(k-1)}{2}(kJ_1+J_2+J_3+J_4)-2I_0+o(1)\right), \tag{5.16}$$

where J_r is given above and

$$I_0 = \int_0^1 P(1-t)^2t^{k-1}dt. \tag{5.17}$$

Therefore given a polynomial P we can get an asymptotic lower bound for S by explicitly calculating the integrals I_0, J_1, J_2, J_3 and J_4.

Explicitly we have for $r=1$

$$J_1 = \int_0^1\tilde{P}(1-t)^2t^{k-2}dt. \tag{5.18}$$

Similarly for $r=2$ we have

$$J_2 = J_{21}+J_{22}+O(\epsilon), \tag{5.19}$$

where

$$J_{21} = \int_0^1\frac{B}{x(B-x)}\int_0^{1-x}\left(\tilde{P}(1-t)-\tilde{P}(1-t-x)\right)^2t^{k-2}dtdx, \tag{5.20}$$

$$J_{22} = \int_0^1\frac{B}{x(B-x)}\int_{1-x}^1\tilde{P}(1-t)^2t^{k-2}dtdx. \tag{5.21}$$

Similarly for $r = 3$ we have

$$J_3 = J_{31} + J_{32} + J_{33} + J_{34} + O(\epsilon), \tag{5.22}$$

where

$$J_{31} = \int_0^{1/2}\int_x^{1-x} \frac{B}{xy(B-x-y)} \int_{1-x}^{1} \left(\tilde{P}(1-t)\right)^2 t^{k-2} dtdydx, \tag{5.23}$$

$$J_{32} = \int_0^{1/2}\int_x^{1-x} \frac{B}{xy(B-x-y)} \int_{1-y}^{1-x} \left(\tilde{P}(1-t) - \tilde{P}(1-t-x)\right)^2 t^{k-2} dtdydx, \tag{5.24}$$

$$J_{33} = \int_0^{1/2}\int_x^{1-x} \frac{B}{xy(B-x-y)} \int_{1-x-y}^{1-y} \left(\tilde{P}(1-t) - \tilde{P}(1-t-x) - \tilde{P}(1-t-y)\right)^2 t^{k-2} dtdydx, \tag{5.25}$$

$$J_{34} = \int_0^{1/2}\int_x^{1-x} \frac{B}{xy(B-x-y)} \int_{0}^{1-x-y} \left(\tilde{P}(1-t) - \tilde{P}(1-t-x) - \tilde{P}(1-t-y) + \tilde{P}(1-t-x-y)\right)^2 t^{k-2} dtdydx. \tag{5.26}$$

Finally for $r = 4$ we have

$$J_4 = J_{41} + J_{42} + J_{43} + J_{44} + J_{45} + J_{46} + J_{47} + J_{48} + J_{49} + J_{410} + J_{411} + O(\epsilon), \tag{5.27}$$

where

$$J_{41} = \int_0^{1/3}\int_x^{(1-x)/2}\int_y^{1-x-y} \frac{B}{xyz(B-x-y-z)} \int_{1-x}^{1} \left(\tilde{P}(1-t)\right)^2 t^{k-2} dtdzdydz, \tag{5.28}$$

$$J_{42} = \int_0^{1/3}\int_x^{(1-x)/2}\int_y^{1-x-y} \frac{B}{xyz(B-x-y-z)} \int_{1-y}^{1-x} \left(\tilde{P}(1-t) - \tilde{P}(1-t-x)\right)^2 t^{k-2} dtdzdydz, \tag{5.29}$$

$$J_{43} = \int_0^{1/4}\int_x^{1/2-x}\int_{x+y}^{1-x-y} \frac{B}{xyz(B-x-y-z)} \int_{1-x-y}^{1-y} \left(\tilde{P}(1-t) - \tilde{P}(1-t-x) - \tilde{P}(1-t-y)\right)^2 t^{k-2} dtdzdydz, \tag{5.30}$$

$$J_{44} = \int_0^{1/3}\int_x^{(1-x)/2}\int_y^{x+y} \frac{B}{xyz(B-x-y-z)} \int_{1-z}^{1-y} \left(\tilde{P}(1-t) - \tilde{P}(1-t-x) - \tilde{P}(1-t-y)\right)^2 t^{k-2} dtdzdydz, \tag{5.31}$$

$$J_{45} = \int_0^{1/4}\int_x^{1/2-x}\int_{x+y}^{1-x-y} \frac{B}{xyz(B-x-y-z)} \int_{1-z}^{1-x-y} \left(\tilde{P}(1-t) - \tilde{P}(1-t-x) - \tilde{P}(1-t-y) + \tilde{P}(1-t-x-y)\right)^2 t^{k-2} dtdzdydz, \tag{5.32}$$

$$J_{46} = \int_0^{1/3}\int_x^{(1-x)/2}\int_y^{x+y} \frac{B}{xyz(B-x-y-z)} \int_{1-x-y}^{1-z} \left(\tilde{P}(1-t) - \tilde{P}(1-t-x) - \tilde{P}(1-t-y) - \tilde{P}(1-t-z)\right)^2 t^{k-2} dtdzdydz, \tag{5.33}$$

$$J_{47} = \int_0^{1/4}\int_x^{1/2-x}\int_{x+y}^{1-x-y} \frac{B}{xyz(B-x-y-z)} \int_{1-x-z}^{1-z}$$
$$\Big(\tilde{P}(1-t) - \tilde{P}(1-t-x) - \tilde{P}(1-t-y)$$

$$- \tilde{P}(1-t-z) + \tilde{P}(1-t-x-y)\Big)^2 t^{k-2} dt dz dy dz, \tag{5.34}$$

$$J_{48} = \int_0^{1/3}\int_x^{(1-x)/2}\int_y^{x+y}\frac{B}{xyz(B-x-y-z)}\int_{1-x-z}^{1-x-y}$$
$$\Big(\tilde{P}(1-t) - \tilde{P}(1-t-x) - \tilde{P}(1-t-y) - \tilde{P}(1-t-z)$$
$$+ \tilde{P}(1-t-x-y)\Big)^2 t^{k-2} dt dz dy dz, \tag{5.35}$$

$$J_{49} = \int_0^{1/3}\int_x^{(1-x)/2}\int_y^{1-x-y}\frac{B}{xyz(B-x-y-z)}\int_{1-y-z}^{1-x-z}$$
$$\Big(\tilde{P}(1-t) - \tilde{P}(1-t-x) - \tilde{P}(1-t-y) - \tilde{P}(1-t-z)$$
$$+ \tilde{P}(1-t-x-y) + \tilde{P}(1-t-x-z)\Big)^2 t^{k-2} dt dz dy dz, \tag{5.36}$$

$$J_{410} = \int_0^{1/3}\int_x^{(1-x)/2}\int_y^{1-x-y}\frac{B}{xyz(B-x-y-z)}\int_{1-x-y-z}^{1-y-z}$$
$$\Big(\tilde{P}(1-t) - \tilde{P}(1-t-x) - \tilde{P}(1-t-y) - \tilde{P}(1-t-z) + \tilde{P}(1-t-x-y)$$
$$+ \tilde{P}(1-t-x-z) + \tilde{P}(1-t-y-z)\Big)^2 t^{k-2} dt dz dy dz, \tag{5.37}$$

$$J_{411} = \int_0^{1/3}\int_x^{(1-x)/2}\int_y^{1-x-y}\frac{B}{xyz(B-x-y-z)}\int_0^{1-x-y-z}$$
$$\Big(\tilde{P}(1-t) - \tilde{P}(1-t-x) - \tilde{P}(1-t-y) - \tilde{P}(1-t-z) + \tilde{P}(1-t-x-y)$$
$$+ \tilde{P}(1-t-x-z) + \tilde{P}(1-t-y-z) - \tilde{P}(1-t-x-y-z)\Big)^2 t^{k-2} dt dz dy dz. \tag{5.38}$$

We choose $k = 22$ and $P(t) = 1 + 60t - 300t^2 + 3500t^3$ and find that

$$I_0 = \frac{121351}{59202} = 2.04978\dots, \tag{5.39}$$
$$J_1 = \frac{228380}{18027009} = 0.01266\dots, \tag{5.40}$$
$$J_2 \geq 0.041 + O(\epsilon), \tag{5.41}$$
$$J_3 \geq 0.048 + O(\epsilon), \tag{5.42}$$
$$J_4 \geq 0.028 + O(\epsilon). \tag{5.43}$$

Thus we have that

$$S \geq \frac{\mathfrak{S}(\mathcal{L})(\log R)^k}{(k-1)!}\left(\frac{0.99(k-1)}{2}(kJ_1 + J_2 + J_3 + J_4) - 2I_0 + O(\epsilon) + o(1)\right)$$
$$\geq \frac{\mathfrak{S}(\mathcal{L})(\log R)^k}{(k-1)!}\left(0.013 + O(\epsilon) + o(1)\right). \tag{5.44}$$

In particular, for N sufficiently large and ϵ sufficiently small we have $S > 0$, and so there are infinitely many n for which an admissible 22-tuple attains at least two prime values and one value with at most 4 prime factors.

The set $\{0, 6, 8, 14, 18, 20, 24, 30, 36, 38, 44, 48, 50, 56, 60, 66, 74, 78, 80, 84, 86, 90\}$ is an admissible 22-tuple, and so the interval $[n, n+90]$ infinitely often contains at least two primes and an integer with at most 4 prime factors.

We remark here that if we can take the level of distribution $\theta = 1 - \delta$ for every $\delta > 0$ then we can take $k = 19$ instead of 22, which reduces the length of the interval to 80.

6. Acknowledgment

I would like to thank my supervisor, Prof. Heath-Brown, for suggesting this problem and for his careful reading of this paper.

References

[1] Elliott, P. D. T. A., and Halberstam, H. A conjecture in prime number theory. In *Symposia Mathematica, Vol. IV (INDAM, Rome, 1968/69)*. Academic Press, London, 1970, pp. 59–72.
[2] Friedlander, J., and Granville, A. Limitations to the equi-distribution of primes. I. *Ann. of Math. (2) 129*, 2 (1989), 363–382.
[3] Goldston, D. A., Graham, S. W., Pintz, J., and Yildirim, C. Y. Small gaps between products of two primes. *Proc. Lond. Math. Soc. (3) 98*, 3 (2009), 741–774.
[4] Goldston, D. A., Pintz, J., and Yildirim, C. Y. Primes in tuples. I. *Ann. of Math. (2) 170*, 2 (2009), 819–862.
[5] Goldston, D. A., Pintz, J., and Yildirim, C. Y. Primes in tuples. II. *Acta Math. 204*, 1 (2010), 1–47.
[6] Halberstam, H., and Richert, H.-E. *Sieve methods*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], London-New York, 1974. London Mathematical Society Monographs, No. 4.
[7] Heath-Brown, D. R. Almost-prime *k*-tuples. *Mathematika 44*, 2 (1997), 245–266.
[8] Maynard, J. Almost-prime *k*-tuples. Pre-print, available at http://arxiv.org/abs/1205.4610v1.
[9] Thorne, F. Bounded gaps between products of primes with applications to ideal class groups and elliptic curves. *Int. Math. Res. Not. IMRN*, 5 (2008), Art. ID rnm 156, 41.

Mathematical Institute, 24-29 St Giles', Oxford, OX1 3LB

E-mail address: maynard@math.ox.ac.uk